

SECURITY AND COMMUNICATION NETWORKS
Security Comm. Networks 00: 1–20 (2009)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sec.0000



Real-time support for HCCA function in IEEE 802.11e networks: a performance evaluation[†]

Gabriele Cecchetti, Anna Lina Ruscelli

ReTis Lab., Scuola Superiore S. Anna, Pisa, Italy

Summary

The IEEE 802.11 standard for wireless networks has been recently enhanced with the IEEE 802.11e amendment which introduces Quality of Service support. It provides differentiation mechanisms at the Medium Access Control layer, using two additional access functions: the Enhanced Distributed Channel Access (EDCA) function and the HCF Controlled Channel Access (HCCA) function. Only the HCCA mechanism is suitable for serving traffic streams with real-time requirements such as multimedia applications and Voice Over IP. The IEEE 802.11e standard does not specify a mandatory HCCA scheduling algorithm, but it offers a reference scheduler as the guideline in the resources scheduling design.

In this paper we analyze four HCCA alternative schedulers to the reference one. They offer real-time guarantees proposing different solutions to the request of QoS and real-time support expressed by the increasing diffusion of multimedia applications. A performance evaluation is conducted to show the main differences between the considered schedulers, including the reference one.

The results show that under several scenarios there is not a unique best scheduler, but there exists a variety of solutions depending on the specified requirements. The conclusions of the paper offer some guidelines in the choice of the scheduler tailored for a particular scenario of interest.

Copyright © 2009 John Wiley & Sons, Ltd.

KEY WORDS: quality of service; real-time guarantees; scheduling algorithms; performance evaluation; wireless lan

1. Introduction

In wireless communications the *Quality of Service* (QoS) and real-time guarantees provisioning are important issues due to the diffusion of mobile devices which support multimedia applications

such as *Voice over IP* (VoIP), multimedia video, and videoconferencing. Furthermore, *Constant Bit Rate* (CBR) and *Variable Bit Rate* (VBR) traffic with different requirements need differentiated QoS levels with real-time guarantees.

On the other hand the space and time-varying characteristics of the wireless channel [1] affect the network performance in terms of *Signal to Interference plus Noise Ratio* (SINR) and *Bit Error Rate* (BER). Thus it is not possible to assure hard QoS constraints in terms of exact values of network and application parameters,

*Correspondence to: Gabriele Cecchetti, Retis Lab., Scuola Superiore S. Anna, Pisa, Italy.
E-mail: g.cecchetti@sssup.it

[†]This paper is an extended version of “Performance Evaluation of Real-Time Schedulers for HCCA Function in IEEE 802.11e Wireless Networks”, presented at ACM Q2SWinet, Vancouver, Canada, October 2008.

including delay, delay jitter, packet loss ratio, and throughput. However the characteristics of multimedia applications allow us to consider soft requirements expressed by admitted intervals for the parameters' values. In fact, for multimedia applications, missing deadlines implies only a degradation of the received QoS and not catastrophic events. Therefore we can consider soft-real time constraints with the exception that we cannot employ the classical methodologies used in static real-time systems because they cannot handle the dynamic characteristics of both the wireless medium and the multimedia applications.

Even if IEEE 802.11 [2] is the recognized standard for Wireless Local Area Networks (WLAN), its earlier releases were designed to provide only best effort services thus it has been recently enhanced with the IEEE 802.11e amendment in order to include the QoS guarantees.

The recent standard has introduced a differentiation mechanism at the *Medium Access Control* (MAC) layer, using two additional access functions: the *Enhanced Distributed Channel Access* (EDCA) function and the *HCF Controlled Channel Access* (HCCA) function. The EDCA function is based on distributed control and enables prioritized channel access, while the latter requires centralized scheduling and allows the applications to negotiate parameterized service guarantees. In particular, since only the HCCA mechanism is suitable for respecting real-time constraints, we consider only this MAC function.

However, the IEEE 802.11e standard does not specify a mandatory HCCA scheduling algorithm, but it offers a reference scheduler, which is compatible with the use of link adaptation, and respects a minimum set of performance requirements. In particular it periodically assigns a fixed transmission time interval to all the managed *Traffic Streams* (TSs).

Some studies evaluated the new standard through analytical techniques and simulations and demonstrated that the HCCA improves the QoS support particularly for CBR traffic. However, it performs poorly with VBR traffic, since it assigns fixed transmission parameters therefore it cannot accommodate the traffic variability. As a consequence several researchers have suggested alternative scheduling algorithms to the reference one, in order to improve its QoS provisioning for

VBR traffic, but they do not consider the real-time constraints specifically.

In this paper we present a performance evaluation of five schedulers tailored for the real-time guarantees support over IEEE 802.11e HCCA networks, in terms of the admission control, the resource utilization efficiency and the access delay. This is neither a fully comprehensive evaluation nor a survey of the existent scheduling algorithms for IEEE 802.11e WLANs, but the analysis of some known algorithms that, to the best of our knowledge, can provide real-time features for WiFi networks.

The rest of the paper is organized as follows: in section 2 we describe the IEEE 802.11 standard, in section 3 we summarize some scheduling algorithms for HCCA function, in section 4 we illustrate the considered schedulers including the reference one and in Section 6 we evaluate their performance. Finally, the conclusions are drawn in Section 7.

2. IEEE 802.11 MAC Protocols

This section provides a description of the IEEE 802.11b and IEEE 802.11e MAC protocols.

2.1. IEEE 802.11b MAC Protocol

The IEEE 802.11b MAC supports two data packet transmission modes. One mandatory, the *Distributed Coordination Function* (DCF) and one optional, the *Point Coordination Function* (PCF).

DCF deals on distributed medium access based on *Carrier Sense Multiple Access/Collision Avoidance* (CSMA/CA), where each node competes with the other ones for access to the medium. The collisions are avoided through a random *backoff* procedure which ensures different waiting transmission times to the stations. Considering that all stations have the same channel access parameters, the same medium access priority, and there is no streams differentiation but the medium access is managed only by means of a random procedure, DCF does not provide QoS support but supplies only best effort service.

The basic feature of PCF is the polling mechanism, managed by the *Access Point* (AP), which allows only polled stations to transmit. During the *Contention Free Period* (CFP), the AP polls its associated stations according to the

polling list, usually in a round-robin manner. If the CFP terminates before all stations have been polled, the polling list will be resumed at the next CFP from the previous stopping point. Since the polling list is fixed and each polled station (STA) occupies the medium until it has data to send, independently from the presence of pollable STAs with more stringent QoS and real-time requirements, PCF is not suitable to provide any form of QoS support.

2.2. The IEEE 802.11e MAC Protocol

The IEEE 802.11e standard has introduced a service differentiation mechanism at the MAC layer. It uses the two additional access functions, EDCA and HCCA, multiplexed by the new *Hybrid Coordination Function* (HCF).

The services differentiation enables the protocol to distinguish the packets with different service requirements, overcoming the limits of the best effort model and delivering data with different QoS levels and negotiated services. The design approach has been to enhance the preexistent MAC functions, DCF and PCF, through the new mentioned ones, which are based on the same corresponding medium access methods. Moreover, in order to ensure compatibility with legacy devices, the standard allows the coexistence of DCF and PCF with EDCA and HCCA.

2.2.1. EDCA

The EDCA function enhances the distributed control of DCF by means of a prioritized channel access which ensures different prioritized QoS levels. The incoming traffic in a STA is classified in four *Access Categories* (AC), corresponding to four different service levels. Each AC has its own transmission queue and its own set of channel access parameters, used for the packets classification. The most important ones are *Contention Window*, which sets the backoff interval, and *Transmission Opportunity* (TXOP) which is the maximum duration of a node transmission. The access to the medium is regulated by two contention phases. The first one, internal to each STA, is won by the AC with the smallest backoff time. This value is used by each STA in the external contention for the wireless medium. So nodes with higher priority can access the channel earlier than the other ones.

2.2.2. HCCA

HCCA provides a centralized polling scheme to enhance PCF with QoS support through a traffic streams classification, based on stations requests, into eight TSs. Each TS can be uni-directional (uplink or downlink) or bi-directional (both of them) and corresponds to a specific service level identified by particular values of protocol parameters. The QoS-aware *Hybrid Coordinator* (HC), usually located at the *QoS Access Point* (QAP), manages the access to the medium.

In order to be included in the polling list of the HC, a *QoS Station* (QSTA) uses the QoS management frame *Add Traffic Stream* (ADDTTS) to send to the QAP a QoS reservation request for each of its TSs. TS parameters are collected in a *Traffic Specification* (TSPEC), negotiated between QSTA and QAP. Mandatory fields are illustrated in table I.

Table I. Traffic specification mandatory fields.

TSPEC parameters	Symbol
Mean data rate (bps)	R
Nominal SDU size (bytes)	L
Minimum PHY rate (bps)	Γ
Delay bound (s)	D
Maximum service interval (s)	MSI

The negotiation of TSPECs parameters during the reservation request allows HCCA to guarantee TSs a parameterized QoS access to the medium.

Moreover, during this phase HC performs the admission control, checking if the acceptance of the requesting TS can compromise the service guarantees of the already admitted TSs. If the TS is admitted, QAP sends to the QSTA a positive acknowledgement which contains also the service start time for the frames of the considered TS. As result of this negotiation HC has to also compute, as aggregation of the QSTA TSPECs, the following transmission parameters sent to the QSTA at the polling time:

Service Interval (SI) : the time interval between two successive polls of the node. It is a submultiple of the 802.11e beacon interval duration

Transmission Opportunity (TXOP) : the node transmission duration based on the mean application data rates of its TSs.

SI and TXOP are the basic parameters used by the scheduling algorithms to manage the access to the medium and their choice is the key in the QoS provisioning.

When there are admitted QSTAs, the QAP listens to the medium and, if it is idle for a *PCF Interframe Space* (PIFS), it gains control of the channel. The QAP, during the *Controlled Access Phase* (CAP), polls a single QSTA at turn, according to the polling list generated by the scheduler considering the QoS and real-time requirements. It is necessary to distinguish between downlink TXOP, during which the QAP sends bursts of QoS Data to QSTA and uplink TXOP, that starts when the polled QSTA takes the medium control. If the polled QSTA does not have packets for the considered TS, i.e. the TS is not backlogged, or if the head-of-line packet does not fit into the remaining TXOP duration, the QSTA sends a QoS CF-Null frame to the QAP. This situation is shown in fig. 1.

3. Related Works

Many research studies evaluated the performance of the reference IEEE 802.11e scheduler employing analytical techniques and simulations [3, 4, 5]. They demonstrated the usefulness of the proposed mechanism, but they also highlighted its limits. In fact, scheduling each QSTA with a fixed couple of transmission parameters, SI and TXOP, allows to provide some relevant results in terms of real-time and QoS support only for CBR traffic, which shows stable and invariant values of its parameters. But in the case of VBR traffic, (like VoIP, videostreaming etc.), a fixed scheduler gives a non-optimal resources utilization since it does not react to the network and traffic variability. The same conclusions are valid for the *Admission Control* which is missing of the necessary flexibility that gives a more stringent condition with the admission of less TSs than possible, wasting the available resources.

Several alternative scheduling algorithms have been proposed to improve the QoS provisioning of IEEE 802.11e networks, in particular in the case of VBR traffic, [6, 7, 8, 3]. Nevertheless they do not consider real-time constraints and few works evaluate the real-time issues of the reference scheduler, i.e. its ability to respect timing constraints expressed in term of TSs deadlines, and propose possible solutions.

Looking at the proposed real-time scheduling algorithms for IEEE 802.11e networks, to the best of our knowledge, they are characterized by the different mechanisms used to provide temporal guarantees. Table II illustrates such mechanisms adopted by some algorithms. The

Table II. Scheduling mechanisms.

Scheduling mechanisms	Algorithms
Queue length estimation	<i>FHCF</i> , ARROW
Feedback based mechanism	FBDS
Timed token based mechanism	<i>WTTP</i>
Deadline-based scheduling	SETT-EDD, <i>RTH</i> , <i>WCBS</i>

algorithms named in the table using italic font will be described more accurately and compared in section 4.

One component of the packet delay is the queue delay, due to the lateness in the transmission queue delivering. A resources assignment tailored to the network traffic can reduce the waiting time experienced by the packets in the transmission queues.

The algorithms can employ different types of information, depending on the use of a theoretical model or the actual values of the queues length. In the latter case, a better resource reservation is possible with respect to it being based on estimated queue length where the assigned TXOP could be major (waste time), minor (delayed packet) or equal (ideal case) with respect to the actual needed transmission time.

In *ARROW* [9] the next TXOP is computed considering the actual buffered TSs data at the beginning of the polling and communicated to the QAP through the QS header field. Moreover, since each new packet could be delayed at least one SI, the scheduler ensures that the deadline is not exceeded and the delay requirements respected bounding the MSI. However the increasing polling overhead, due to the fact that the MSI upper limit could be less than half of the standard, is compensated by a more accurate TXOP computation. Knowing the queue length for each TS, the scheduler can manage differently the TSs with different requirements. Finally the next QSTA to be polled is chosen by means of an *Earliest Due Date* (EDD) [10] deadlines scheduling.

The *Application-Aware Adaptive HCCA Scheduler* [11], derived from ARROW, distinguishes

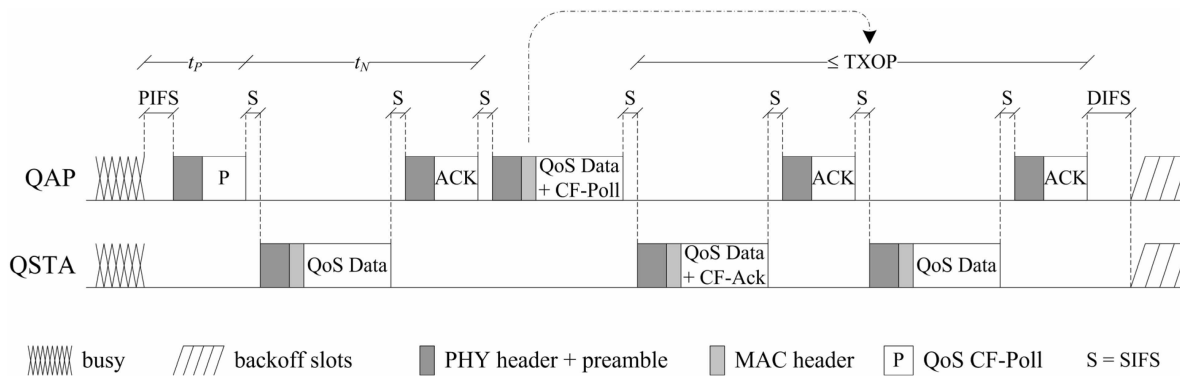


Figure 1. Sample HCCA exchange sequence.

uplink and downlink schedulers, whereas the *Earliest Deadline First* (EDF) algorithm [12] adapts the polling order to the stations' requirements, expressed by means of the computed deadlines. The uplink scheduler assigns each QSTA a minimum and a maximum SI, modified to follow the application and network conditions and to digest the buffered traffic, respecting the QoS requirements. They are computed considering the mean TSPEC values and depending of periodic or aperiodic traffic.

A feedback mechanism can be used to update the value of transmission parameters to the network variability. All the QSTAs and their transmission queues are regarded as a system whose balance is perturbed by new incoming TSs. The *Feedback Based Dynamic Scheduler* (FBDS) [13] behaves as a closed loop controller which restores this balance by bandwidth recovering, limiting the maximum delay. It assigns dynamically, by means of a feedback mechanism, the TXOP according to queue length estimation at beginning of the new CAP phase through a discrete time model, while SI remains fixed. Moreover it compensates the errors produced by channel perturbations not previewed by the estimation algorithm using the actual queue length information sent by each QSTA.

In [7] the authors propose the *Scheduling Estimated Transmission Time - Earliest Due Date* (SETT-EDD) algorithm, which is based on the use of the deadline concept. The algorithm uses variable TXOP and SI but it maintains the same admission control as the reference scheduler. Both the TXOPs and the SIs values are limited

between the minimum and a maximum SI so that transmissions can be done at least at PHY rate thus respecting the deadlines. The polling order is managed by the real-time algorithm *Delay-Earliest Due Date* (Delay-EDD) [14], which polls the QSTA according to the non decreasing deadlines.

4. The Analyzed Real-Time HCCA Schedulers

In this section we analyze with more detail the schedulers chosen for our performance evaluation. We begin describing the limits of the IEEE 802.11e HCCA reference scheduler in terms of real-time provided guarantees. Then we describe with more detail the schedulers chosen for our performance evaluation as tailored for the real-time provisioning for HCCA WLANs.

The scheduling algorithm and the admission control test of each scheduler are analyzed. In particular the admission control test represents a sufficient feasibility condition for the scheduling algorithm and it allows for determination of the set of streams which can transmit within their timing constraints. When a new TS asks the QAP for the right to be transmitted, the scheduler checks if the new traffic stream can be admitted to the medium without jeopardizing the guarantees of already admitted streams and if it would require more capacity than the system can provide. If the TS cannot be admitted the QSTA is notified of insufficient available capacity, otherwise the scheduler updates the total used bandwidth and allocates data structures to perform the TS scheduling.

4.1. The Reference Scheduler

The reference scheduler proposed by the IEEE 802.11e standard suggests how to compute the main protocol parameters, SI and $TXOP$, suitable to meet the requirements globally expressed by each QSTA. Different values for each specific QSTA are computed only for $TXOP$, whereas SI is computed as a unique value for all non-AP QSTAs with admitted streams.

4.1.1. Reference Scheduler Algorithm

The goal to assure that each QSTA transmits its TSs during an interval tailored to its requirements is solved considering the mean values of transmission parameters.

The SI is suggested to be less than the beacon interval, ensuring that the QSTAs will be polled at least one time during the beacon duration. It has also to be less than the minimum of the *Maximum SI* (MSI_i) of each $QSTA_i$. This means that the QSTAs are polled within an interval less than the minimum of those requested for each of their TSs and this assures no deadline misses. For that which concerns the MSI_i , the standard suggests to compute it as the ratio between the *Delay Bound* and the possible number of retries. The standard states that the number of retries may be chosen to meet a particular probability of dropping a packet when it exceeds its Delay Bound.

The $TXOP$ is globally assigned to a QSTA and not to its single TSs. It is computed using the SI previously obtained and the mandatory negotiated TSPEC parameters. First of all the scheduler computes the max number N_i of max size *MAC Service Data Unit* (MSDU) transmitted at the mean data rate R_i :

$$N_i = \left\lceil \frac{SI \cdot R_i}{L_i} \right\rceil$$

Then it computes $TXOP$ as the maximum between the time to transmit N_i MSDU with nominal size L_i and the time to transmit one max sized MSDU (M_i , i.e. 2304 bytes) at the Physical Transmission Rate (R_i), which is equal to the minimum PHY rate Γ_i of the $QSTA_i$:

$$TXOP_i = \max\left(\frac{N_i \cdot L_i}{\Gamma_i}, \frac{M_i}{\Gamma_i}\right) + O$$

This choice is conservative, since it preserves the maximum time to transmit, at the minimum PHY

rate, the maximum number of bytes that can arrive during SI . Note that $TXOP_i$ takes into account the transmission overhead (O) in time units: such overhead includes interframe spaces, ACKs and CF-Polls.

SI and $TXOP_i$ are fixed values, based on “worst case” conditions, and they are recomputed only if a new TS arrives with a MSI_i greater than the preexistent ones. This means that all the QSTAs are polled with the same period, SI , and that the different TSs of a station i are served with the same computation time, $TXOP_i$.

4.1.2. Reference Scheduler Admission Control

The admission control test suggested for the *Admission Control Unit* (ACU) of the reference scheduler depends on the values of SI and $TXOP$ computed for each TS as described before. It assures that the ratio between the $TXOP$ and the SI of the new admitted TS, added to the sum of the ratio of $TXOP_i$ and SI of each already admitted TS_i , must not exceed the portion of bandwidth reserved to the HCCA function:

$$\frac{TXOP_{k+1}}{SI} + \sum_{i=0}^k \frac{TXOP_i}{SI} \leq \frac{T - T_{CP}}{T} \quad (1)$$

where k is the number of already admitted streams, $k + 1$ is the index of the newly arriving stream, T is the beacon interval and T_{CP} is the time used for EDCA traffic.

4.2. Fair HCF

Fair HCF (FHCF) [15] aims to improve the fairness of both CBR and VBR traffic and the delay performances assigning variable $TXOP$ s by means of the estimation of the uplink TSs queues length.

The mathematical model proposed for the queues shows the relationship between polling interval and queuing delays and it is used to estimate the global packet delay. More specifically, it distinguishes between the packet queuing delay Q and the waiting time delay W . The former is due to the delay in the queue, influenced by the variations in packet size and data rate, while the latter is defined as the interval between the packet arrival time and the QSTA polling time. The authors distinguish the case when the queues are empty at the end of the $TXOP$, especially for

CBR traffic, and the case where the queues are not empty, which is more realistic for real wireless networks. In the first situation, there is a further distinction if the packet arrives before the polling time of its QSTA (the packet has to wait for the polling QSTA):

$$d_i(t) = \left(\sum_{j=1}^{i-1} T_j - t + \frac{q_i M_i}{R_{eff}} + \frac{\rho_i}{R_{eff}} \right)$$

or if the packet arrives during (the packet has to wait for the transmission of previous queued packets) or after (the packet has to wait for the next SI):

$$d_i(t) = \left(SI - t + \sum_{j=1}^{i-1} T_j + \frac{\rho_i}{R_{eff}} \left(t - \sum_{j=1}^{i-1} T_j \right) \right)$$

where T_j is the allocated TXOP for TS_j , q_i is the number of packets in the queue, M_i is the MSDU size, R_{eff} is the effective data throughput, ρ_i is the application data rate. In this case the delay is bounded by the SI. Instead, if the queues are not empty at the TXOP end, the packets can be queued later than the next SI since the delays are cumulative. In the best case the packet will be transmitted in the next SI but in general, if there is highly variable traffic, the packet delay can become really unpredictable. This delay is expressed by:

$$D_i = \max_t \{d_i(t)\} = SI - T_i + \frac{q_i^e M_i}{\rho_i}$$

where q_i^e is the queue length at the end of the TXOP T_i .

These equations suggest two different ways to control the maximum delay: increase the TXOP T_i or reduce the SI but this implies an increasing number of polling and an increasing overhead. FHCF adopts the alternative method to reduce the delay through the control of the queue length before the polling time. To deal with these delays the authors designed FHCF with two schedulers: the QAP scheduler and the node scheduler.

The QAP scheduler estimates the varying queue length for each QSTA at the beginning of the next SI, q_i^{est} , and compares this value with the ideal one, q_i^{ideal} . In particular,

$$q_i^{est} = \frac{\rho_i(SI - t_i^e)}{M_i} + q_i^e$$

where t_i^e is the time when q_i^e is evaluated. Since sending rate and packet size can change, this estimation can not be accurate, so the QAP scheduler corrects the q_i^e computation using its expected value. Then it computes the additional required time t_i^{est} (positive or negative) for each TS of a QSTA and reallocates the corresponding TXOP duration according to the number of additional packets in the queue. Moreover, it evaluates the actual available time after the allocation of all the TXOPs in one SI in the ideal case and, if it is not sufficient, it decreases fairly all the assigned times by a percentage of t_i^e . This allows management of the traffic variations.

The node scheduler, located in each QSTA, just after the CF-Poll reception, can redistribute the additional time of the TXOP, which is always globally allocated to the QSTA, among its TSs. It executes the same computation than the QAP scheduler but more accurately since each QSTA knows exactly its TSs queues size at the beginning of the polling and it is able to estimate its queue length at the end of TXOP and the requested TS additional time. According to its allocated TXOP, it evaluates the remaining time T' that can be re-allocated considering the number of packets N_i to transmit in the TS_i and the time required to transmit a packet, computed according to its QoS requirements:

$$T' = T - \sum_{i=1}^p N_i \cdot \left(\frac{M_i}{R_{eff}} + 2SIFS + ACK \right).$$

FHCF uses the same admission control test of the reference scheduler (eq. 1).

4.3. Wireless Timed Token Protocol

Wireless Timed Token Protocol (WTTT) [16] is based on the *Timed Token Protocol* (TTP) [17], a token passing MAC protocol for ring-based networks. The token is used to manage a circular list of nodes in a round-robin manner. Each node refers downlink and uplink TSs. One special node represents contention traffic using contention-based schemas as EDCA and DCF.

4.3.1. WTTT Scheduler Algorithm

The scheduler visits each node of the list for a time called *sojourn time* and either schedules TXOP or refrains from generating CAP according

to the node TSs, (if the node represents contention traffic). The nodes are inserted in the round robin list only if they actually have traffic to be served:

- *downlink TSs* are added and removed whenever they become backlogged or idle;
- *uplink TSs* are added and removed if each QSTA *piggybacks* the backlog of its TSs on outgoing data messages (standard IEEE 802.11e feature). However such nodes are put back in the round robin list at least each minimum SI.

The token circulation is ruled by the *Target Token Revolution Time* (TTRT), a protocol parameter selected as reference round duration which corresponds to the SI: its value is computed by the QAP according to the TSPEC values negotiated by the QSTA during the admission control phase. The authors set the value of TTRT to half the smallest delay bound, since the round robin duration is bounded by $2 \cdot TTRT$:

$$TTRT = \frac{1}{2} \min_i \{D_i\}.$$

This maximum limit makes the scheduler more conservative in terms of maximum delay tolerable by the QSTAs, implying that some QSTAs can be polled more frequently than necessary, increasing the system overhead.

The sojourn time is given by either one or both the following two components:

- *synchronous bandwidth*: a fixed time H_i computed as a percentage of TTRT and used to transmit frames with HCCA rate based guarantees
- *asynchronous bandwidth*: a variable and not reserved portion of TTRT used to transmit the remaining not guaranteed traffic.

In other words, each node i has a $H_i \geq 0$ and a *Token Rotation Timer* (TRT_i), initially set to TTRT, which counts down the time from the last server visit to obtain the maximum fair share of *asynchronous bandwidth* that node i can exploit. When a node is served, the asynchronous bandwidth is computed as follows:

$$a_i = \begin{cases} 0 & TRT_i < 0, \\ \min\{TTRT - H_i, TRT_i\} & TRT_i \geq 0. \end{cases}$$

In this way the asynchronous TSs are not reserved any capacity: they may transmit frames only if

the token arrives earlier than expected, i.e., before a TTRT time has elapsed from the last token visit. The early arrival of the token usually occurs when the synchronous TSs consumed less than the reserved capacity during the previous token revolution. The assignment of the sojourn time, which is substantially the TXOP, is different by the other approach. In fact in the case of QoS traffic H_i is computed as a fixed fraction of TTRT: this means that the time destined to the QoS TSs is not recomputed and, in case of unused resources, the remaining time is not assigned to other QSTAs but to the non-QoS traffic of the polled QSTA. Since the contention traffic is managed by the EDCA function this implies that the QSTA has to change the medium access mode adopting the contention method and, when it has exhausted the recovered time, it has to go back to the HCCA. Obviously, this increases the system overhead.

4.3.2. WTP Admission Control

Since $TTRT$ is the average inter-service time for a node, and $2 \cdot TTRT$ is an upper bound the admission control test for the WTP algorithm is the following:

$$\sum_{node_i} H_i + \tau \leq TTRT. \quad (2)$$

Therefore a node that has a synchronous bandwidth equal to H_i is in fact entitled to an average rate equal to $H_i/TTRT$ times the channel speed, and has a bounded medium access time. The term τ is an overhead due to the time required to regain the control of the medium and to start a new CAP after a DCF/EDCA phase.

4.4. Real-Time HCCA

The *Real-Time HCCA* (RTH) algorithm [18] is designed to provide real-time support in HCCA assuring the traffic streams a fixed amount of capacity during a fixed period.

4.4.1. RTH Scheduler Algorithm

The periodic scheduler, based on EDF algorithm plus *Stack Resource Policy* (SRP) [19], takes into account the non-preemptability of frame transmissions, that are considered critical sections. The scheduler activity is split into *offline* activity at stream lifetime timescale, which performs the

more complex activity, and *online* activity at the frame transmission timescale.

Since the scheduling parameters are computed offline, the online activity consists only of reading the next entry $[i, t_i, \text{TXOP}_i]$, composed by the index of the next QSTA which can access the medium, the polling time t_i and the duration of its transmission. Conversely, the offline activity attends to the admission control and to the timetable computation. The admission control phase takes into account the requirements of the new and admitted QSTAs and provides the parameters used in the timetable computation.

When the *MAC sublayer Management Entity* has admitted a QSTA it translates a set of the TSPEC mandatory parameters $[R_i, N_i, D_i, \Gamma_i]$ into that used in the enforcement procedure and notifies the QSTA of successful operation. Each TS is characterized through two parameters, the *capacity* C_i and the *period* T_i , derived from the TSPEC set. The capacity C_i is computed as the time needed to send at rate R_i how many Nominal SDU can be held during a period T_i , and it can be expressed as follows:

$$C_i = \left\lceil \frac{R_i \cdot T_i}{N_i} \right\rceil \cdot t_{N_i}$$

where t_{N_i} is the *Nominal transmission time*. The period T_i is:

$$T_i = \begin{cases} D_i & D_i < \left\lceil \frac{N_i}{R_i} \right\rceil, \\ \left\lceil \frac{R_i}{N_i} \cdot D_i \right\rceil \cdot \frac{N_i}{R_i} & \text{otherwise.} \end{cases}$$

4.4.2. RTH Admission Control

In order to admit a TS the QAP verifies the schedulability test introduced in the multi-programmed environment by SRP which includes the blocking time due to the non-preemptability of TS transmission. In the case of HCCA, the minimum critical section b_i for a TS_i is equal to the nominal size SDU transmission time t_{N_i} including the poll time for uplink TSs, t_{P_i} , i.e., $b_i = t_{N_i} + t_{P_i}$.

When TS_i is in a critical section and is scheduled instead of the highest priority TS_j , TS_i is said to block TS_j . So, the blocking time for a TS_i is the maximum critical section durations of TSs with a period longer than TS_i :

$$B_i = \max_{j > i} \{b_j\}.$$

The schedulability analysis produces the following sufficient condition to determine the set of n schedulable TSs:

$$\frac{B_i}{T_i} + \sum_{j \leq i} \frac{C_j + \pi_j \cdot t_{P_j}}{T_j} \leq 1 \quad \forall i : 1 \leq i \leq n. \quad (3)$$

where π_j is the maximum number of times that the QAP has to poll TS_j , during T_j . Moreover, larger versions of B_i and b_i , \overline{B}_i and \overline{b}_i , which are the maximum quantities that the equality holds in the equation, are used in order to reduce the MAC overhead.

4.5. Wireless Capacity Based Scheduler

The *Wireless Capacity Based Scheduler* (WCBS) for HCCA [20], suitable for serving Soft Real-Time applications, is derived from the basic ideas of the Constant Bandwidth Server [21], a soft real-time scheduling algorithm for real-time operating systems. The original algorithm is based on EDF and provides a mechanism to serve multimedia applications with soft real-time constraints along with hard real-time applications served by hard real-time algorithms.

In WCBS, the basic idea is that WiFi networks handle traffic streams instead of tasks: some of them require temporal guarantees, while others are just best effort traffic streams. The nature of the wireless medium does not allow transmissions with hard real-time requirements, so here we do not need a different real-time algorithm for them, different to the operating systems. The best effort traffic is served during the contention phase, while the traffic streams requiring QoS guarantees will be served during the contention-free phase according to the HCCA protocol.

4.5.1. WCBS Static and Dynamic Parameters

WCBS uses static and dynamic parameters to rule the transmitted packet scheduling.

The scheduler assigns to each TS_i an ordered pair of static parameters:

Q_i : the budget, i.e., the maximum transmission time which can be assigned during a period;

P_i : the service interval of the TS_i .

In particular, Q_i is the maximum capacity, expressed in time units, that a stream i can consume in its period P_i . These parameters are

computed during the admission control phase and their values are based on the $TSPEC_i$. They do not change during normal conditions. The ratio $U_i = Q_i/P_i$ is denoted as the *factor utilization* of the stream, i.e., the TS_i bandwidth.

During the scheduling, each TS_i is characterized by the following dynamic parameters, which represent the actual stream status:

c_i : the current capacity, i.e., the remaining time that can be assigned to TS_i during the next TXOP;

d_i : the absolute deadline before the budget transmission time has to finish;

p_i : the next time an uplink TS_i will be polled when it has no more data to transfer or it has exhausted its TXOP;

state : the current state of the stream can be one between *transmitting*, *active*, *polling*, *idle*.

The scheduling process marks each TS with a status according to its condition and its parameters' values. It is used to determine the subsequent temporal evolution and it can be:

transmitting if the TS is transmitting packets;

active if the TS is in the transmitting queue because it has a packet to send and $c_i > 0$. Moreover if it is an uplink stream it is the next TS which will be polled;

idle if the TS is a downlink stream that has no packet to transmit or has exhausted its capacity;

polling if the TS is an uplink stream and it is in the polling queue because it has packets to send and $c_i > 0$ but is still too early to be polled.

Note that only one stream at time can be in the transmitting state.

4.5.2. WCBS Admission Control

The admission control test of WCBS is:

$$\sum_{i=0}^N \frac{Q_i}{P_i} \leq \frac{T - T_{CP}}{T} \quad (4)$$

where P_i is computed as the maximum SI and Q_i is determined by means of a weighted function f of

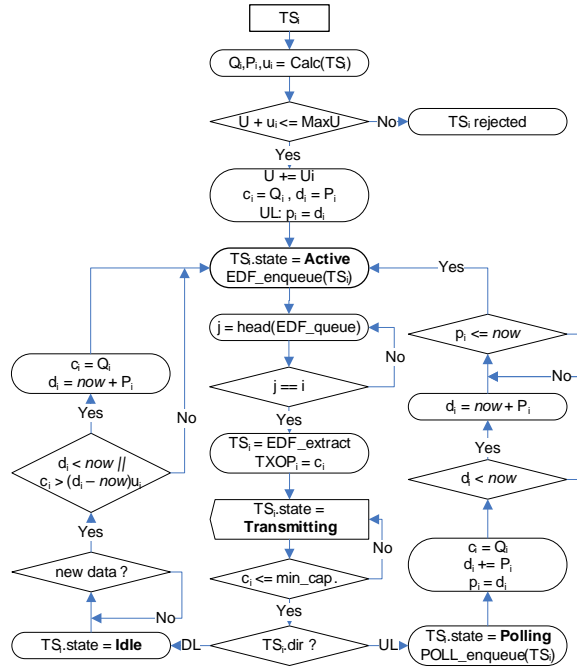


Figure 2. WCBS scheduling algorithm.

Q_{min} and Q_{max} , which evaluates the minimum and maximum budget needed to transmit during a period T_i , respectively, Nominal SDUs at the mean data rate and Maximum SDUs at the peak data rate. Q_{min} and Q_{max} are expressed as:

$$Q_{min} := \left\lceil \frac{R_i \cdot P_i}{L_i} \right\rceil \cdot t_N, \quad Q_{max} := \left\lceil \frac{\Lambda_i \cdot T_i}{ML_i} \right\rceil \cdot t_N$$

where P_i is the period, R_i is the mean data rate, Λ_i is the peak data rate, L_i is the nominal SDU size, ML_i is the maximum SDU size for the i^{th} TSPEC, and t_N is the Nominal transmission time computed as:

$$t_N = t_{DATA} + t_{SIFS} + t_{ACK} + t_{SIFS} \\ = L_i \cdot \Gamma_{DATA} + h_{ACK} \cdot \Gamma_{ACK} + 2 \cdot t_{SIFS} \quad .$$

4.5.3. WCBS Scheduling Algorithm

After the admission control phase, the temporal evolution of the scheduler is as follows (see Fig.2).

1. For each new admitted TS_i , at the beginning: $c_i = Q_i$, $d_i = now + P_i$, $state = active$, where *now* is the current time.
2. Whenever a TS_i is **active** its transmission request is enqueued in an EDF queue.

3. The streams are *served* in EDF order: the scheduler extracts the next TS_i to serve from the top of EDF queue, and it sets: TS_i state to **transmitting**, $TXOP = c_i$, then it decreases the capacity c_i by the effective transmission time.
4. When a TS_i finishes transmitting, the next pending transmission, if any remains, is served using its current capacity and deadline.
5. When a TS_i with $c_i < \text{min_capacity}^\dagger$ is served, if TS_i is an uplink stream, its state becomes **polling**, it is inserted in the polling queue, and the following quantities are set: $c_i = Q_i$, $d_i = d_i + P_i$, $p_i = d_i$. If its deadline is still expired, i.e., $d_i < \text{now}$, then it is postponed to another period by now, i.e., $d_i = \text{now} + P_i$. In this way, using this *recharging mechanism*, a TS_i does not have to wait for a deadline expiration to recharge its capacity and then it is ready earlier to transmit again.
6. A TS_i remains in the **polling** state until $p_i \leq \text{now}$, then it is extracted from the polling queue, it becomes **active** and it is inserted in EDF queue.
7. When an **idle** downlink TS_i is served because it has new data to transmit, if $c_i \geq (d_i - \text{now})U_i$ the scheduler recharges the stream capacity to the maximum value, $c_i = Q_i$, and it generates a new deadline by a period from now: $c_i = Q_i$, $d_i = \text{now} + P_i$. Then it becomes **active** and it is inserted in EDF queue.
8. If there are no active streams a *Contention Period* is started.

5. Scheduling Algorithms Properties Analysis

In this section we briefly analyze the selected algorithms from a theoretical and mathematical point of view. This general dissertation does not have the aim to provide a complete analytical study but a theoretical validation of the performance evaluation which is the goal of

[†] min_capacity is the minimum capacity needed to transmit an SDU, and eventually CF-Poll for an uplink TS_i .

the present work. So we have chosen to highlight the temporal isolation property, which assures a protection mechanism in the transmission, and the computational complexity, which evaluates the mathematical efficiency of the selected algorithms.

5.1. Temporal Isolation

The use of the standard TXOP parameter in order to assign a maximum transmission time to each admitted QSTA is the key in the provisioning of negotiated QoS levels. Moreover the TXOP parameter is functional to provide the temporal isolation property to the scheduling algorithms.

In particular for WCBS the temporal isolation property is derived from the analogue one of the Constant Bandwidth Server algorithm. The introduction of the budget for each traffic stream provides a solution to execution overruns, which happen when a TS asks to be transmitted more than expected, jeopardizing the temporal guarantees of other streams. The budget assignment performs a bandwidth reservation: each stream is assigned a fixed capacity, and when it requires additional time, its deadline is postponed. In this case such streams might experience a delay, whereas the guarantees assured for the other streams remain unaffected.

The same conclusions are valid for FHCF and RTH. Moreover, RTH considers the blocking time derived by the granularity of the TSs transmission that implies a minimum length of assured transmission duration.

From this point of view WTTP is more conservative, again. In fact its QoS TXOPs are fixed and the unused bandwidth is recovered for the contention traffic. Thus, if a QSTA consumes less than expected the non-QoS traffic will be sent, but this does not jeopardize the QoS traffic that has been already sent.

5.2. Computational Complexity

The evaluation of the computational complexity of the considered algorithms highlights their efficiency in term of computational resource utilization.

WCBS can be split into two phases, each having a particular task and a related complexity. The first is the insertion of the admitted TS in the right place within the EDF queue, following a deadline-based order: the operation of inserting an

element into an ordered queue of n elements, the n admitted streams, has a complexity $O(\log n)$. Furthermore, if we consider the worst case where n TSs are admitted at the same time, the complexity grows to $O(n \cdot \log n)$. The second phase, which involves the extraction of the stream from the queue, has a complexity of $O(1)$.

In RTH the admission control test has an $O(n)$ computational complexity. Admitted TSs are scheduled following the timetable computation based on EDF-order of periods T_i , for $i = 1..n$. Its computational complexity is due to the selection of the TS to be transmitted between that with unfulfilled capacity, so it is equal to $O(n)$ in function of number n of admitted TSs. Instead the computational complexity of the online scheduler activity is $O(1)$ in terms of the number of QSTAs.

The same considerations can be applied to the reference and the FHCF schedulers. In fact, after the admission control test, which is the same for both, and the computation of the transmission parameters they do not apply any particular algorithm in the polling list sorting but accept that which is performed during the negotiation phase and expressed in terms of QoS requirements. Since the sorting algorithm is not specified, we can infer that it can be characterized, in general, by an $O(n)$ complexity due to the comparison of requested QoS levels, while the extraction of an element from the polling list has an $O(1)$ complexity.

The WTTP computational complexity is bounded by the algorithm which enforces the polling order. In fact the Timed Token Protocol manages the token circulation between the nodes imposing the sequence of polled QSTAs. So the QAP does not have to perform any other selection or ordination activity. The simplicity gives a $O(1)$ complexity. This is the drawback between the adoption of a simple solution with a predetermined ordination algorithm, which limits its computational complexity, and the lack in the flexibility.

6. Performance Analysis

In this section we present some results of the algorithms performance evaluation obtained through simulation. We first describe the simulation settings and the used traffic model. Then we discuss the results about admission control analysis, the null rate experienced over

the medium, the unreserved capacity available for contention traffic and the mean access delay.

As shown by the obtained results an optimal algorithm suitable to meet all the different requirements as a unique solution does not exist. In fact each requirement involves different parameters sometimes in opposition. Tuning some of them to obtain the desired scheduler behavior from a particular point of view as specified by particular values of network parameters can imply the degradation of other performance. Therefore the choice of the global optimal scheduler is bounded by the drawback of opposite trends of the parameters and is obtained through the use of a specific algorithm tailored for the desired behavior.

6.1. Simulation Settings

We used the physical layer parameters specified by the High Rate Direct Sequence Spread Spectrum (HR-DSSS) (Table III).

Table III. MAC/PHY simulation parameters

Parameter	Value	Parameter	Value
SIFS	10 μs	PHY header	192 μs
PIFS	30 μs	Data rate	11 Mb/s
DIFS	50 μs	Basic rate	1 Mb/s
SlotTime	20 μs	Bit error rate	0 b/s

MAC level fragmentation, multirate support, RTS/CTS protection mechanism are disabled and we assume that all nodes can directly communicate with each other, without the hidden node problem. We use the HCCA implementation described in [22] for the *ns-2* network simulator [23], as a framework to implement the proposed algorithm.

The analysis has been carried out using the method of independent replications. Specifically we ran independent replications of 600 seconds each with 100 seconds warm-up periods until the 95% confidence interval is reached for each performance measure. Confidence intervals are not drawn whenever negligible.

6.2. Traffic Model

We use two types of uplink (UL) traffic streams requiring QoS guarantees: VoIP and video. The VoIP traffic is simulated using a VoIP generator module for *ns-2* described in [24].

The VoIP streams of packets are modeled as an ON/OFF source: during the ON (*talkspurt*) periods the traffic is CBR with parameters that depend on the encoding scheme; during the OFF (*silence*) periods no packets are generated. Talkspurt and silence periods are distributed according to the Weibull distribution [25] that models a one-to-one conversation: $\lambda_{ON}=1.423s$, $k_{ON}=0.824s$, $\lambda_{OFF}=0.899s$, $k_{OFF}=1.089s$ (which yields $E[ON]=1.58s$, $E[OFF]=0.87s$). The employed encoding schemes are G.711, G.723.1 and G.729A [26] with the parameters as shown in Table IV. For both encoding schemes we set the TSPEC delay bound to the packet interarrival time (*period*) and the mean data rate to the peak rate during talkspurts.

Table IV. VoIP encoding schemes

Codec	G711	G723.1	G729A
Frame size (s)	80	30	10
Period (s)	0.02	0.0455	0.02
Sample per packet	2	1	2
Payload size (B)	160	30	20
IP/UDP/RTP			
Header size (B)	40	40	40
SDU size (B)	200	70	60
Data rate (b/s)	80000	12320	24000

The video stream traffic is generated using pre-encoded MPEG4 trace files from the Internet archive of traces [27]. An MPEG4 encoder produces streams of variable size frames at fixed intervals [28]. They are chosen to represent a videoconference session (*LectureHQ-Reisslein* trace file) and a video streamed over the network (*Jurassic Park High Quality* trace file). The TSPEC parameters are shown in table V.

Table V. Traffic parameters for video streams

video stream	VideoConf.	VideoStr.
Mean frame size (B)	660	3800
Max frame size (B)	11386	16745
Period (s)	0.033333	0.040
Mean data rate (b/s)	157712	770000
Peak data rate (b/s)	2732640	3300000

Best effort data traffic is transmitted using legacy DCF. Stations with data traffic operate in asymptotic conditions, i.e., they always have a frame to transmit. The packet length of data traffic is constant and equal to 1500 bytes.

6.3. Admission Control Analysis

We evaluated the number of admitted stations using the considered schedulers under different scenarios: CBR traffic only, VBR traffic only and mixed traffic. This approach shows the schedulers behavior under different traffic conditions. An analytical comparison of the different admission control formulas confirms and explains the simulation results.

Fig. 3 shows the number of admitted G.729A TSs as a function of admitted G.711 streams. Since they are CBR TSs and their codecs have

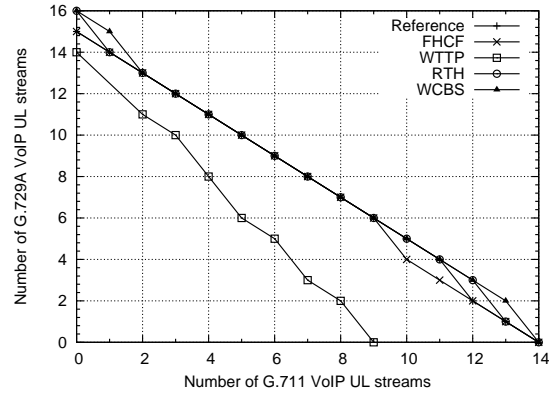


Figure 3. Admission Control: number of admitted VoIP G.729A vs. G.711 UL TSs.

the same SI, the number of admitted TSs is substantially similar for each scheduler except for WTTT which underutilizes the medium. The poor performance of WTTT is due to the fact that, as explained previously, in the case of QoS traffic the sojourn time H_i is computed as a fixed fraction of TTRT and the unused resources are destined to the non-QoS traffic of the polled QSTA. This simulation shows that the considered schedulers, except WTTT, behave similarly in the case of CBR traffic.

In Fig. 4 we analyze the number of admitted G.723.1 TSs as a function of admitted G.711 streams. Since G.723.1 can have two different rates, here we show that the reference scheduler and FHCF perform worse. This is due to the codecs having different periods (20 ms and 45.5 ms for G.711 and G.723.1, respectively) while these schedulers poll the stations at the smallest interarrival period (thus 20ms). This is often more than needed and the computed TXOP for TSs is overestimated. RTH and WCBS have better performance because they are derived by EDF so

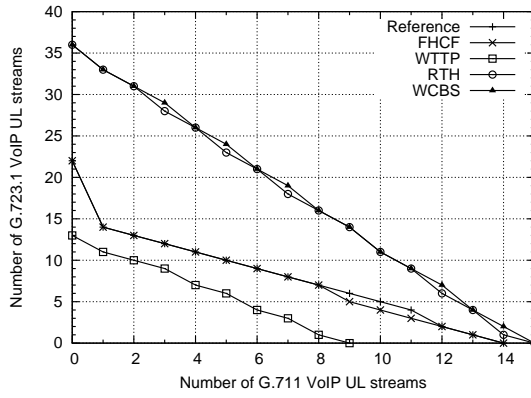


Figure 4. Admission Control: number of admitted VoIP G.723.1 vs. G.711 UL TSs.

they allow a more precise estimation of the needed computation time for variable traffic. They assign different SI and TXOP for each TSs. The WTTP admission control performs even worse than the reference scheduler because it takes into account only an *a priori* evaluation of the lowest Delay Bound of considered TSs. Such a parameter is the most strict requirement with respect to the other admission control tests. This simulation highlights that the reference scheduler lacks in the flexibility required for the variable traffic.

Finally, Fig. 5 shows the number of admitted TSs under different scenarios involving two different kinds of VBR streams, the videoconference and the videostreaming, as a function of the number of admitted VoIP G.711 and G.729A TSs. These TSs have four different periods: 33ms, 40ms,

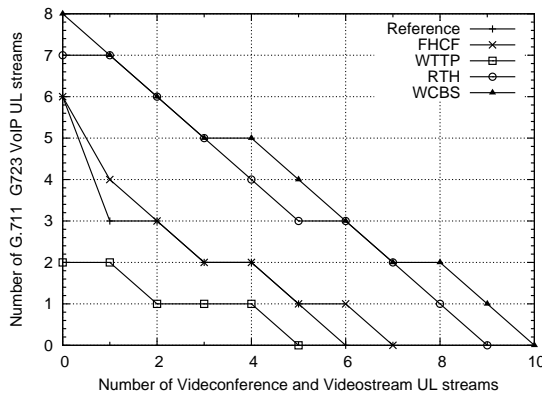


Figure 5. Admission Control: number of admitted VoIP G.711 and G.723.1 vs. VC and VS UL TSs.

20ms and 45.5ms. In this scenario the schedulers

able to deal with such different TSPECs are winning. In fact WCBS and RTH perform better than the others, and in particular WCBS admits more TSs because it computes a smaller TXOP and is able to recover the unused bandwidth with more flexibility. On the other hand, the reference scheduler cannot efficiently manage TSs with different TSPECs, because it polls TSs with different periods more frequently than needed, setting the scheduling duration to the smallest TS period and assigning an overestimated TXOP to the TSs. Similarly, WTTP performs even worse because of its pessimistic admission control test.

To make an analytical evaluation of the different admission control tests, now we put their equations in the same form and use the same notation:

$$Refer.: \frac{T_{CP}}{T} + \sum_{i=0}^k \frac{TXOP_i}{SI} \leq 1 \quad (5)$$

$$FHCF: \frac{T_{CP}}{T} + \sum_{i=0}^k \frac{TXOP_i}{SI} \leq 1 \quad (6)$$

$$WTTP: \frac{\alpha}{TTRT} + \frac{\sum_{node_i} H_i + \tau}{TTRT} \leq 1 \quad (7)$$

$$RTH: \frac{T_{CP}}{T} + \frac{B_i}{T_i} + \sum_{j \leq i} \frac{TXOP_j}{T_i} \leq 1 \quad (8)$$

$$WCBS: \frac{T_{CP}}{T} + \sum_{i=0}^k \frac{TXOP_i}{T_i} \leq 1 \quad (9)$$

where the $TXOP_i$ comprises the tp_i polling time for uplink streams and the RTH admission control test takes into account the contention period. Note that:

- eq. 5 and eq. 6 are the same by definition;
- eq. 7 is composed of one term representing the asynchronous traffic and one term representing the synchronous traffic. If we set α so that $\alpha/TTRT = T_{CP}/T$, we have to compare only the second term which represents the HCCA portion of the traffic. Because $TTRT = \frac{1}{2} \min_i \{D_i\}$ and, in general, the SI is chosen to not make the $\min_i \{D_i\}$ expired, we can approximate as $TTRT = \frac{1}{2} \cdot SI$. Since this quantity is smaller than SI, the ratio of WTTP is greater than the ratio of the reference scheduler, so the index of the sum can reach a smaller value, admitting less QSTAs;

- eq. 8 and eq. 9 differ only for the term B_i/T_i which takes into account the critical section involved in SRP and has a small value because usually $T_i \gg B_i$; for this reason WCBS and RTH always admit almost the same number of TSs;
- eq. 9 differs from eq. 5 only by the second term; because in general $T_i \geq SI$ so the ratio $TXOP_i/T_i \leq TXOP_i/SI$ and therefore the index of the sum can reach a higher value.

6.4. Efficiency Analysis

This section is dedicated to analyzing the schedulers' efficiency under different scenarios. The scheduler efficiency is evaluated as a measure of how well it utilizes the network resources. Such evaluation is done through the analysis of the null rate and the polling interval experienced during the polling of the QSTAs and considering the throughput left to DCF and EDCF TSs.

The null rate is defined as the number of *Null* packets received by the QAP after it has sent a *CF-Poll* frame: this happens when a QSTA has no packets to transmit. Evaluating the null rate, we can check if the polling time computation is suitable for the considered traffic or that the QAP is polling the QSTAs more frequently than necessary, increasing the system's overhead.

We consider a scenario with four VoIP G.711 uplink TSs and four VoIP G.723.1 uplink TSs. In Fig. 6 we show the null rate value for the analyzed schedulers. With G.711 TSs the null rate

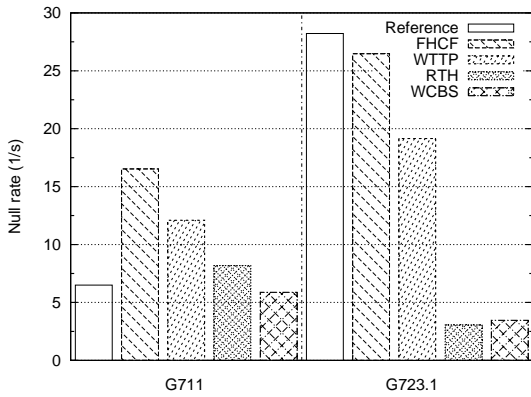


Figure 6. Null rate of UL VoIP TSs.

is almost the same for the reference, RTH and WCBS because they have the same SI and the same computed TXOP and this type of traffic

does not need more scheduling flexibility. The null rate is higher for FHCF and WTTP, because both try to empty the queue node polling more aggressively. With G.723.1 TSs the reference, WTTP and FHCF schedulers show a higher null rate because they use the shortest polling period of 20 ms instead of 45.5 ms. WCBS and RTH poll the TSs only at their packet arrival, so their polling mechanism is more efficient. In particular, FHCF performs worse with respect to the other schedulers since it adapts the TXOP to the traffic variations but maintains the same SI. This means that even if the scheduler exhausts the queued packet until the next SI, which is a good choice, it polls the QSTA when it does not have any packet to send. Note that *Null* messages are only due to the silence periods, and they are unavoidable.

In fig. 7 we show the polling intervals for a scenario with a mix of VoIP and videoconference TSs. The reference and the FHCF schedulers compute the same polling interval, even if there are TSs with different interarrival time. Instead the WCBS and RTH calculate two different values which allow polling the QSTAs when they have packets to send. The particular behavior of WTTP is due to the fact that it initially computes for the considered QSTA a fixed polling interval $TTRT_i$ equal to TTRT, and only during the transmission does it try to follow the traffic variability by changing this value. But this adaptation is performed slowly as shown in the figure, due to its poor reactivity. This simulation highlights the

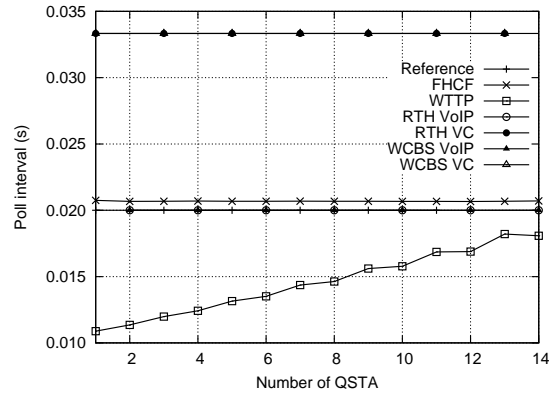


Figure 7. Poll interval of mix of VoIP and videoconference UL TSs.

scheduling efficiency in adapting or not adapting the SI parameter to the traffic variability. The

choice of the polling interval affects also the system overhead.

Now we analyze the unreserved capacity that the schedulers leave for contention-based traffic, scheduled by DCF or EDCF functions.

In fig. 8 we show the unreserved (E)DCF capacity as a function of the increasing number of QSTAs with VoIP and VBR traffic. Obviously,

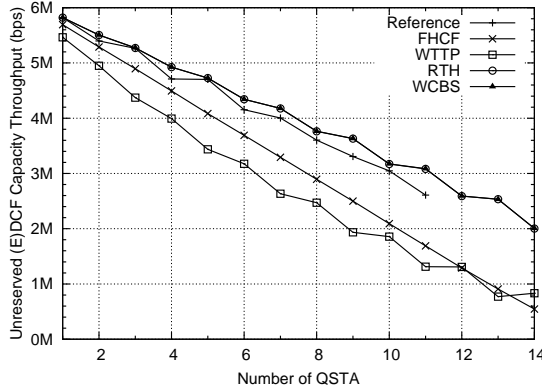


Figure 8. Unreserved DCF capacity throughput with VoIP and VBR traffic.

as the number of QSTAs increases the capacity left for the best effort traffic becomes poor since the available resources are used to serve the QoS traffic. Therefore the schedulers present different levels of efficiency. The reference and WTTP schedulers perform worse since the QSTAs TXOPs (H_i for WTTP) are fixed. In particular WTTP computes TXOP and SI considering an a priori evaluation of the lowest Delay Bound, as mentioned in the admission control analysis, so its admission control is more stringent, the resource management is less efficient and the system overhead is higher. These factors impact in the poor resources left to the non-QSTAs. FHCF, which adapts the TXOPs to the traffic variability considering the transmission queues length, shows a better behavior. It efficiently manages the available resources, leaving more capacity to the best effort traffic. The flexibility of RTH and WCBS in the TXOP and SI computation allows a more efficient admission control and less polling overhead that's why they show the best performance, managing efficiently the available QoS capacity leaving more resources for the non-QSTAs.

It is interesting to analyze the schedulers' behavior with or without ideal channel conditions.

Thus we considered the presence of a uniform channel error. In fig. 9 we compare the unreserved (E)DCF capacity throughput when there are 10 QSTAs transmitting using VoIP codec G.729A. As expected, the schedulers' performance decreases. In fact all the considered schedulers do not implement any error recovery strategy, and the computed TXOPs, even if they are different for the various TSs, are not recomputed after the admission control and negotiation phase. Thus the schedulers are not able to recover the channel error. In particular, only WTTP has a minor loss in the unreserved capacity since it polls the QSTAs more aggressively, "compensating" the non-ideal conditions due to the channel error. So the QSTAs are able to react to the losses by taking advantage of the more frequent polling to retransmit the corrupted packets.

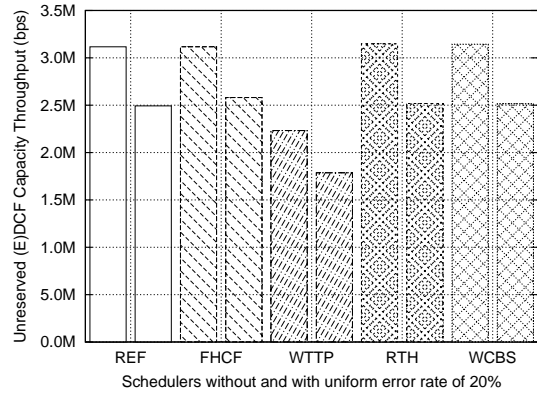


Figure 9. Unreserved (E)DCF capacity throughput when 10 QSTAs are transmitting using VoIP G.729A without or with uniform error rate.

Finally, we consider the QSTAs throughput with the different schedulers. Fig. 10 shows the throughput of the last admitted QSTA when there are 10 admitted stations with G.729A traffic, in the presence or the absence of channel errors. Since this is not a saturation condition and the traffic is CBR, the QSTA has almost the same throughput with the different schedulers. In fact we know that they perform similarly in case of CBR traffic. Only WTTP exhibits lower performance due to the more frequent polling. When introducing the uniform channel error, as inferred from fig. 9, the QSTA experiences a drop in the throughput. The behavior of one single QSTA highlights the previous considerations

about the schedulers, that are missing an error recovery policy.

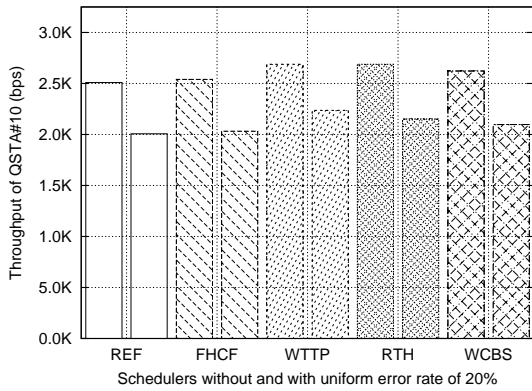


Figure 10. Throughput of QSTA #10 with VoIP G.729A TS without or with uniform error rate.

Since the schedulers do not implement any error compensation mechanism it is not meaningful to analyze their behavior with non-homogeneous channel error. In fact as with uniform error the performance drops down, we can easily induce that the inhomogeneity in the channel conditions impacts severely in the schedulers' activity. Furthermore, no particular scheduler is suitable to avoid this effect.

6.5. Delay Analysis

In this section we investigate the access delay defined as the time elapsed from the packet reaching the MAC layer to that of the packet being successfully acknowledged.

Fig. 11 shows that, in case of CBR traffic like that produced by G.729A codec, all the analyzed schedulers have almost the same mean delay, as expected. The only exception is WTTP: the lower polling interval assigns the QSTAs a sooner opportunity to transmit the TSs, reducing the experienced delay. Note that the delay value is reported only for admitted QSTA. Looking at the *Cumulative Distribution Function* (CDF) of the access delay shown in fig. 12, we have a clear confirmation that the results about the mean delay are correct. Even this analysis shows that WTTP has higher probability to experience a smaller value of access delay while the others perform almost with the same probability for the same values with a little advantage for the reference scheduler.

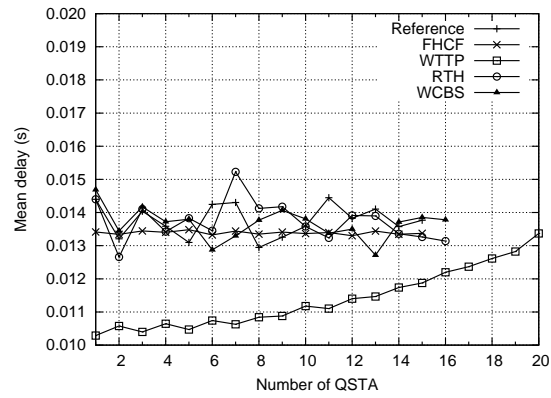


Figure 11. Mean delay of G.729A uplink TSs.

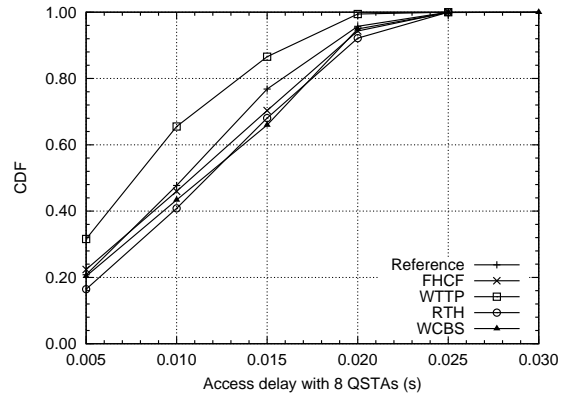


Figure 12. CDF of access delay of 8 QSTAs transmitting VoIP G.729A TSs.

In Fig. 13 we consider a scenario with an increasing number of upload videoconference TSs. We note that the EDF-based schedulers produce

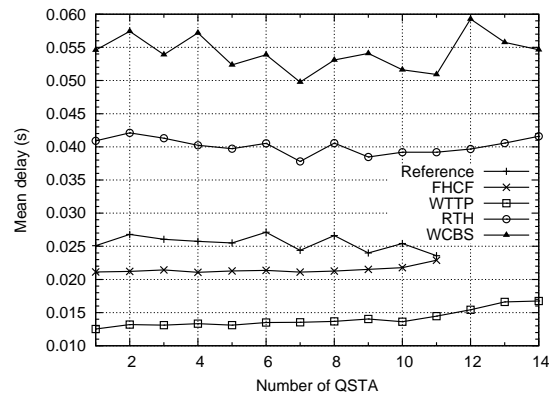


Figure 13. Mean delay of VC uplink TSs.

a mean delay greater than the others. In fact

the EDF algorithm executes a new sorting for each CAP phase, while the other schedulers maintain a fixed order of TSs. WCBS performs even worse because it also postpones the current deadline when there is not enough capacity to transmit. This motivates its shown high variability in the mean delay too. The reference and FHCF schedulers perform similarly because FHCF changes only the TXOP parameter while keeping SI fixed. WTTP has better behavior since it assigns the TTRT, corresponding to the SI, a value smaller than the minimum tolerable delay, ensuring that all the QSTAs respect their timing requirements. So WTTP is more conservative in minimizing the experienced delay at the cost of an increased system overhead, due to the reduced polling interval, as we explained in the previous section.

Observing the CDF of the access delay experienced by a QSTA in the same scenario, when 8 QSTAs are transmitting we see that both RTH and WCBS cause the VBR traffic to have a lower probability keeping low access delay than the other schedulers. FHCF and WTTP are the best for this kind of traffic.

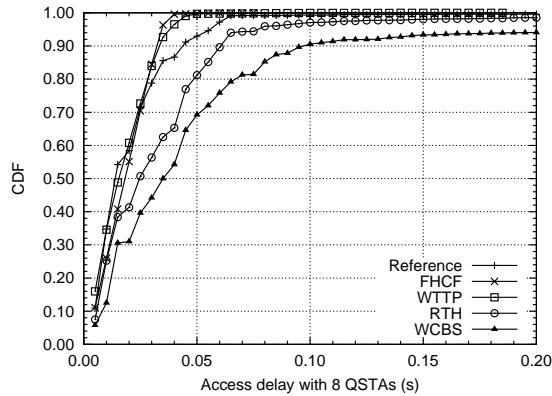


Figure 14. CDF of access delay of 8 QSTAs transmitting VC TSs.

Now we want to take into account how the access delay changes in the presence of a uniform error rate over the wireless channel. In this scenario every packet has a probability of 0.25 to be lost.

In fig. 15 the mean access delay experienced by each scheduler under this scenario is shown. Note that now every scheduler is affected by the loss of packets. While in fig. 13 the mean values span the range from 0.01 s to 0.06 s, now this range

is increased to span the interval from 0.02 s to 0.25 s. In particular WCBS and RTH mean access delays increase more than 3 times: this is greater than the Delay Bound admitted for that TSPEC. In such a case the packet will be discarded by the receiving client. Only FHCF and WTTP still have an acceptable value, even if it is increased.

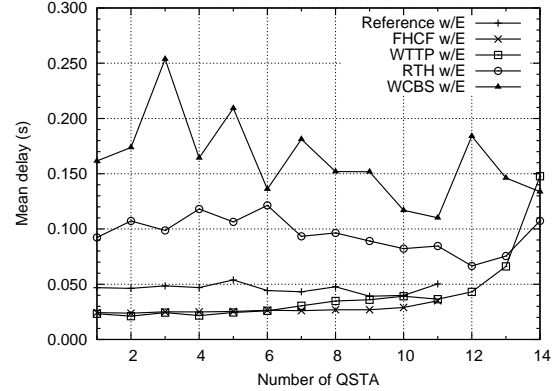


Figure 15. Mean delay of VC uplink TSs in the presence of uniform error rate over the wireless channel.

The effect produced by the uniform error rate on the access delay is visible even if we look at its cumulative distribution function shown in fig. 16. Comparing this figure with fig. 14 we note that the slope of each probability curve is decreased, meaning that every scheduler has less probability to keep the access delay under a specific value.

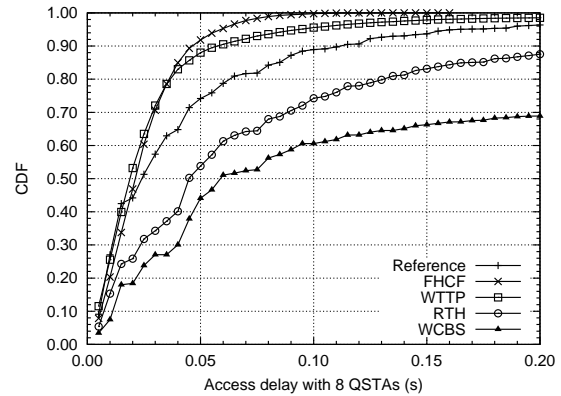


Figure 16. CDF of access delay of 8 QSTAs transmitting VC TSs in the presence of uniform error rate over the wireless channel.

As explained in the throughput analysis, since the evaluated schedulers do not use any type

of error recovery, the channel error directly impacts the experienced performance, without any reaction by the schedulers. All the schedulers show very poor performance with a considerable increase in the mean delay. The distinctions between the different algorithms are the same as in the case of no error. Thus FHCF and WTTP perform better, whereas the poor ones are the EDF-based, RTH and WCBS. In particular the latter shows the same high variability than in the ideal conditions, due to the postponed deadlines.

7. Conclusions

In this paper we have compared five different schedulers for HCCA IEEE 802.11e networks, namely, reference scheduler, FHCF, WTTP, RTH and WCBS. We have described their parameters, admission control tests, and their temporal evolution.

In order to evaluate how suitable they are to support traffic streams requiring soft real-time guarantees, we have analyzed their characteristics. In particular the admission control phase has been tested under different scenarios both analytically and through simulation. Then the efficient use of the medium has been studied considering the null rate and the polling interval produced with mixed traffic streams. Our results have been confirmed by evaluating the unreserved capacity available for contention-based traffic. While the admission control and the efficiency tests have shown that EDF-based schedulers perform better than the others, the delay analysis with VBR scenarios has illustrated that RTH and WCBS result in greater access delays. The drawback of the reference, FHCF and WTTP schedulers is essentially due to restrictive admission control test, fixed SI, and overestimated TXOP.

There is not a clear winning strategy among the five proposed schedulers: the scheduler for HCCA function has to be chosen according to the scenario. However EDF-based algorithms such as RTH and WCBS seem to be more suitable when applications require temporal guarantees. In particular, WCBS has a simpler design, lower computational complexity, slightly better efficiency than RTH, and it has the feature to postpone the deadlines if required: the increased delay experienced by the TS transmission could be easily reduced using part of EDCA function to transmit further packets or by inserting WCBS in

a more general hierarchical scheduler for different traffic types.

Summarizing the obtained results we can extract some guidelines useful in the choice of the scheduler tailored for a particular scenario of interest. In table VI we classify the behavior of the analyzed schedulers, with respect to different parameters, in four levels of goodness ($-$, $-$, $+$, $++$). We consider the performance of the algorithms in term of computational complexity ($O()$), admission control quality (AC), efficiency and overhead (Eff./Ov.), type of traffic: CBR, mixed of CBR (M-CBR), VBR, mixed traffic (MIX), and finally in term of reactivity to non-ideal channel conditions (Ch.Err).

Table VI. Guidelines for choosing the right scheduler.

	REF	FHCF	WTTP	RTH	WCBS
A.C.	-	-	--	+	++
Eff./Ov.	-	-	--	+	++
$O()$	+	-	+	-	++
CBR	+	+	-	++	++
M-CBR	-	+	-	++	++
VBR	-	+	++	-	-
MIX	-	+	-	-	+
Ch.Err.	-	+	++	-	--

Acknowledgement

This work has been supported by the FRESOR European project (Contract n. 034026).

References

1. Y. L. Boudec and T. Thiran, "A short tutorial on network calculus I: Fundamental bounds in communication networks," in *IEEE ISCAS*, 2000, pp. IV-93-IV-96.
2. IEEE802.11, "Wireless LAN medium access control (MAC) and physical layer (PHY) specification," IEEE, Piscataway, NJ, 2007.
3. S. Tsao, "Extending earliest due-date scheduling algorithms for wireless networks with location dependent errors," in *Proc. IEEE VTC*, Boston, MA, Sept. 2000.
4. A. Grilo and M. Nunes, "Performance evaluation of IEEE 802.11e," in *Proc. PIMRC*, vol. 1, Lisboa, Portugal, Sept. 2002, pp. 511 - 517.
5. J. Cowling and S. Selvakennedy, "A detailed investigation of the IEEE 802.11e HCF reference scheduler for VBR traffic," in *Proc. ICCCN*, Chicago, US, Oct 2004.
6. H. Fattah and C. Leung, "An overview of scheduling algorithms in wireless multimedia networks," *IEEE Wireless*, vol. 9, no. 5, pp. 76-83, Oct. 2002.

7. A. Grilo, M. Macedo, and M. Nunes, "A scheduling algorithm for QoS support in IEEE 802.11e networks," *IEEE Wireless Communications*, vol. 10, no. 3, pp. 36–43, June 2003.
8. S. Lu, V. Bharghavan, and R. Srikant, "Fair scheduling in wireless packet networks," *IEEE/ACM Trans. Net.*, vol. 7, no. 4, pp. 473 – 489, Aug. 1999.
9. D. Skyrianoglou, N. Passas, and A. K. Salkintzis, "Arrow: An efficient traffic scheduling algorithm for ieee 802.11e hcca," *IEEE Trans. Wireless Communications*, vol. 5, no. 12, pp. 3558 – 3567, Dec. 2006.
10. J. R. Jackson, "Scheduling a production line to minimize maximum tardiness," University of California, Los Angeles, CA, Research Report 43, 1955, management Science Research Project.
11. I. Inanc, F. Keceli, and E. Ayanoglu, "An adaptive multimedia qos scheduler for ieee 802.11e wireless lans," in *Proc. IEEE ICC*, Istanbul, Turkey, June 2006.
12. C. Liu and J. Layland, "Scheduling algorithm for multiprogramming in a hard-real-time environment," *Journal of ACM*, vol. 20, no. 1, Jan. 1973.
13. G. Boggia, P. Camarda, L. G. Grieco, and S. Mascolo, "Feedback-based control for providing real-time services with the ieee 802.11e mac," *IEEE/ACM Trans. Net.*, vol. 2, no. 15, pp. 323 – 333, April 2007.
14. D. Ferrari and D. Verma, "A scheme for real-time channel establishment in wide-area networks," *IEEE Journal on Selected Areas in Communications*, vol. 8, no. 3, pp. 368–379, April 1990.
15. P. Ansel, Q. Ni, and T. Turletti, "FHCF: A simple and efficient scheduling scheme for IEEE 802.11e wireless lan," *Mobile Networks and Applications*, vol. 11, no. 3, pp. 391–403, June 2006.
16. C. Cicconetti, L. Lenzini, E. Mingozzi, and G. Stea, "An efficient cross layer scheduler for multimedia traffic in wireless local area networks with IEEE 802.11e HCCA," *ACM Mob. Comput. and Commun. Reviews*, vol. 11, no. 3, pp. 31–46, 2007.
17. R. Grow, "A timed-token protocol for local area networks," in *Proc. Electro/82*, ser. Token Access Protocols, I. Electronic Conventions, Ed., no. Paper 17/3, El Segundo, CA, May 1982.
18. C. Cicconetti, L. Lenzini, E. Mingozzi, and G. Stea, "Design and performance analysis of the real-time hcca scheduler for ieee 802.11e wlans," *Comput. Netw.*, vol. 51, no. 9, pp. 2311–2325, 2007.
19. T. P. Baker, "Stack-based scheduling for real-time processes," *Real-Time Syst.*, vol. 3, no. 1, pp. 67–99, 1991.
20. G. Cecchetti, A. L. Ruscelli, and F. Checconi, "W-cbs: A scheduling algorithm for supporting qos in ieee 802.11e," in *Proc. ACM-IEEE QSHINE*, Vancouver, Canada, Aug. 2007.
21. L. Abeni and G. Buttazzo, "Integrating multimedia applications in hard real-time systems," in *IEEE Real Time-Systems Symposium*, Dec. 1998, pp. 4–13.
22. C. Cicconetti, L. Lenzini, E. Mingozzi, and G. Stea, "A software architecture for simulating IEEE 802.11e HCCA," in *Proc. 3rd IPS MoMe*, Warsaw, Poland, Mar 2005.
23. *Network Simulator* 2, <http://www.isi.edu/nsnam/ns/>.
24. A. Bacioccola, C. Cicconetti, and G. Stea, "User-level performance evaluation of voip using ns-2," in *Proc. NSTools*, 2007.
25. P. T. Brady, "A model for generating on-off speech pattern in two way conversation," *Bell System Technical Journal*, vol. 48, pp. 2445–2472, 1969.
26. *Traffic Analysis for Voice over IP*. Cisco Press, November 2001.
27. <http://traces.eas.asu.edu/>, 2005.
28. F. H. P. Fitzek and M. Reisslein, "Mpeg4 and h.263 video traces for network performance," *IEEE Network*, vol. 15, no. 6, pp. 40–54, November 2001.

Authors' Biographies

Gabriele Cecchetti is a researcher at the Scuola Superiore Sant'Anna University, Pisa, Italy. His current research activities span several areas, including: design and performance evaluation of scheduling algorithms for QoS provisioning in wireless computer networks with soft real-time constraints, design and evaluation of cross-layer framework architecture for providing QoS in heterogeneous wireless networks, and wireless link emulation.

Anna Lina Ruscelli received the MS degree in Electronic Engineering from the University of Perugia, Italy, in 2005. She is currently a PhD student in Innovative Technologies of Information and Communications Engineering and Robotics at Scuola Superiore Sant'Anna, Pisa, Italy. Her research interests include Quality of Service support over heterogeneous networks, IEEE 802.11e networks, and cross-layer design.